

Information-theoretic measures associated with rough set approximations

Ping Zhu^{a,b}, Qiaoyan Wen^b

^a*School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*

^b*State Key Laboratory of Networking and Switching, Beijing University of Posts and Telecommunications, Beijing 100876, China*

Abstract

Although some information-theoretic measures of uncertainty or granularity have been proposed in rough set theory, these measures are only dependent on the underlying partition and the cardinality of the universe, independent of the lower and upper approximations. It seems somewhat unreasonable since the basic idea of rough set theory aims at describing vague concepts by the lower and upper approximations. In this paper, we thus define new information-theoretic entropy and co-entropy functions associated to the partition and the approximations to measure the uncertainty and granularity of an approximation space. After introducing the novel notions of entropy and co-entropy, we then examine their properties. In particular, we discuss the relationship of co-entropies between different universes. The theoretical development is accompanied by illustrative numerical examples.

Keywords: Rough set, entropy, co-entropy, uncertainty, granularity

1. Introduction

To handle inexact, uncertain or vague knowledge in some information systems, Pawlak developed rough set theory in the early 1980s [14, 15]. Since then we have witnessed a systematic, world-wide growth of interest in rough set theory and its applications in a number of fields, such as granular computing, data mining, decision analysis, pattern recognition, and approximate reasoning [12, 17, 18, 30, 34, 35].

The starting point of rough set theory in [14, 15] is the idea that elements of a universe having the same description are indiscernible with respect to the available information. The indiscernibility was described by an equivalence relation in the way that two elements are related by the relation if and only if they are indiscernible from each other. As is well known, any equivalence relation defined on a universe U determines a partition of U into a collection of equivalence classes (blocks): each class contains all and only the elements that are mutually equivalent among them. Any partition π of U represents a piece of knowledge about the elements of U forming a classification and so any equivalence class induced by π is interpreted as a granule of knowledge contained in (or supported by) π .

According to Pawlak's terminology expressed in [16], any subset X of the universe U is called a concept in U . If the concept X is a union of equivalence classes from π , then X is precise in π , otherwise X is vague. The basic idea of rough set theory consists in replacing vague concepts with a pair of precise concepts, its lower and upper approximations [16], and thus, a basic problem in this framework is to reason about the accessible granules of knowledge. To this end, various knowledge granulations (also, information granulations or granulation measures), as an average measure of knowledge granules, have been proposed and addressed in [1, 3, 8, 11, 13, 21, 23, 24, 25, 26, 28, 32]. Among them, there are several information-theoretic measures of uncertainty or granularity for rough sets [1, 3, 8, 10, 11, 13, 21, 23, 25], which are based upon the important notion of entropy introduced by Shannon [22]; for more details, we refer the reader to the excellent survey papers [2, 27].

It is worth noting that the information-theoretic measures mentioned above are only dependent on the sizes of equivalence classes (essentially, the underlying partition) and the cardinality of the universe, independent of the lower

Email addresses: pzhubupt@gmail.com (Ping Zhu), wqy@bupt.edu.cn (Qiaoyan Wen)

and upper approximation operators. For example, in [6, 13, 23, 26] the information entropy $H(\pi)$ of the partition $\pi = \{U_1, U_2, \dots, U_k\}$ is defined as

$$H(\pi) = - \sum_{i=1}^k \frac{n_i}{n} \log \frac{n_i}{n},$$

where n_i is the cardinality of U_i and $n = \sum_{i=1}^k n_i$. As a result, it often yields that some partitions like $\{\{1\}, \{2\}\}$ and $\{\{1, 2\}, \{3, 4\}\}$ have the same entropy (or co-entropy). This seems somewhat unreasonable since the basic idea of rough set theory aims at describing vague concepts by the lower and upper approximations. In other words, the result of this description relies on both the partition and the approximations. In light of this, we should pay more attention to the lower and upper approximation operators.

The previous observation motivates us to propose another information-theoretic entropy function to measure the uncertainty associated to the partition and the approximation operators in this paper. More concretely, given a universe U with n elements and a partition π of U , we take count of the subsets of U described by every pair of lower and upper approximations. Assume that r_i , $1 \leq i \leq m$, is the number of subsets described by the rough set approximation (A_i, A'_i) and every subset of U appears with the same probability. It follows that the rough set approximation (A_i, A'_i) appears with the accumulative probability $r_i/2^n$ since the amount of all subsets of U is precisely 2^n . In this way, we obtain a probability distribution

$$P(\pi) = \left(\frac{r_1}{2^n}, \frac{r_2}{2^n}, \dots, \frac{r_m}{2^n} \right).$$

It gives rise to an information entropy, say $\mathcal{H}(\pi)$, according to Shannon's information theory [22]. On the other hand, we can get by the probability distribution a co-entropy $\mathcal{G}(\pi)$. It turns out that $\mathcal{H}(\pi) + \mathcal{G}(\pi) = n$. After exploring some properties of the entropy and co-entropy, we discuss the relationships of co-entropies between different universes. Roughly speaking, the co-entropy monotonically increases when the partition becomes coarser. For example, the co-entropy of $\{\{1, 2\}, \{3, 4\}\}$ is greater than that of $\{\{1\}, \{2\}\}$.

The remainder of the paper is structured as follows. In Section 2, we briefly review some basics of Pawlak's rough set theory and the information-theoretic measures of uncertainty and granularity for rough sets in the literature. Section 3 is devoted to our novel notions of entropy and co-entropy and their properties. We address the relationship of co-entropies between different universes in Section 4 and conclude the paper in Section 5 with a brief discussion on the future research.

2. Preliminaries

This section consists of two subsections. We briefly recall the definition of Pawlak's rough sets in the first subsection and then review two information-theoretic measures of uncertainty and granularity in rough set theory in the second subsection.

2.1. Rough sets

We start by recalling some basic notions in Pawlak's rough set theory [14, 15].

Let U be a finite and nonempty universal set, and let $R \subseteq U \times U$ be an equivalence relation on U . Denote by U/R the set of all equivalence classes induced by R . Such equivalence classes are also called *elementary sets*; every union (not necessarily nonempty) of elementary sets is called a *definable set*.

For any $X \subseteq U$, one can characterize X by a pair of lower and upper approximations. The *lower approximation* $\underline{app}_R X$ of X is defined as the greatest definable set contained in X , while the *upper approximation* $\overline{app}_R X$ of X is defined as the least definable set containing X . Formally,

$$\underline{app}_R X = \cup \{C \in U/R \mid C \subseteq X\} \text{ and } \overline{app}_R X = \cup \{C \in U/R \mid C \cap X \neq \emptyset\}.$$

The pair $(\underline{app}_R X, \overline{app}_R X)$ is referred to as the *rough set approximation* of X . It follows immediately from definition that $\underline{app}_R X \subseteq X \subseteq \overline{app}_R X$ for any $X \subseteq U$.

The ordered pair $\langle U, R \rangle$ is said to be an *approximation space*. A *rough set* in $\langle U, R \rangle$ is the family of all subsets of U having the same lower and upper approximations. Thus, the general notion of rough set can be simply identified with the rough approximation of any given set.

Recall that a *partition* of U is a collection of nonempty subsets of U such that every element x of U is in exactly one of these subsets; such subsets making up the partition are called *blocks*. We write $\Pi(U)$ for the set of all partitions of U and $\mathcal{P}(U)$ for the power set of U . It is well-known that the notions of partition and equivalence relation are essentially equivalent, that is, for any equivalence relation R on U , the set U/R is a partition of U , and conversely, from any partition π of U , one can define an equivalence relation R_π on U such that $U/R_\pi = \pi$ in the obvious way. Thus, we sometimes say that the ordered pair $\langle U, \pi \rangle$ is an approximation space and write $\underline{app}_\pi X$ and $\overline{app}_\pi X$ for $\underline{app}_{R_\pi} X$ and $\overline{app}_{R_\pi} X$, respectively. More generally, we will use equivalence relation and partition indiscriminately.

If a universe U has more than one element, it is always possible to introduce at least two canonical partitions: One is the trivial partition, denoted by $\check{\pi}$, consisting of a unique block, and the other is the discrete partition, denoted by $\hat{\pi}$, consisting of all singletons from U . Formally,

$$\check{\pi} = \{U\} \text{ and } \hat{\pi} = \{\{x\} \mid x \in U\}.$$

We now define a partial order “ \leq ” on $\Pi(U)$: For any $\pi, \sigma \in \Pi(U)$, $\sigma \leq \pi$ if and only if for any $C \in \sigma$, there exists $D \in \pi$ such that $C \subseteq D$. For instance, $\hat{\pi} \leq \pi \leq \check{\pi}$ for any $\pi \in \Pi(U)$. We say that σ is *finer* than π and that π is *coarser* than σ if $\sigma \leq \pi$. When $\sigma < \pi$, that is, $\sigma \leq \pi$ and $\sigma \neq \pi$, we say that σ is *strictly finer* than π and that π is *strictly coarser* than σ . Informally, this means that σ is a further fragmentation of π .

2.2. Information-theoretic measures

In this subsection, we review two information-theoretic measures associated with rough sets in the literature. These measures are concerned with the uncertainty or granularity of knowledge provided by a partition.

In [6, 13, 23, 26], Shannon entropy [22] has been used as a measure of information for rough set theory as follows. For subsequent need, we fix a notational convention: Throughout the paper, all logarithms are to base 2 unless otherwise specified.

Definition 2.1 ([6, 13, 23, 26]). *Let $\langle U, \pi \rangle$ be an approximation space, where the partition π consists of blocks U_i , $1 \leq i \leq k$, each having cardinality n_i . The information entropy $H(\pi)$ of partition π is defined by*

$$H(\pi) = - \sum_{i=1}^k \frac{n_i}{n} \log \frac{n_i}{n}, \text{ where } n = \sum_{i=1}^k n_i. \quad (1)$$

When $\pi = \check{\pi}$, the entropy function H achieves the minimum value 0, and when $\pi = \hat{\pi}$, it achieves the maximum value $\log n$. Moreover, it has been shown in [23] that for any two partitions π and σ of U , if $\sigma < \pi$, then $H(\sigma) > H(\pi)$.

The equation (1) can be rewritten as follows:

$$H(\pi) = \log n - \sum_{i=1}^k \frac{n_i}{n} \log n_i. \quad (2)$$

Recall that the Hartley measure [7] of uncertainty for a finite set X is

$$H(X) = \log |X|,$$

where “ $|X|$ ” denotes the cardinality of the set X . It measures the amount of uncertainty associated with a finite set of possible alternatives, the nonspecificity inherent in the set.

The first term $\log n$ (i.e., $\log |U|$) in Eq. (2) is exactly the Hartley measure of U , which is a constant independent of any partition. The second term of the equation is basically an expectation of granularity with respect to all blocks in a partition. This quantity has been used by Yao to measure the granularity of a partition in [26] and has been defined by Liang and Shi as the rough entropy of knowledge in an approximation space in [11]. This quantity has also been referred to as co-entropy by some scholars (see, for example, [2, 3]).

Definition 2.2 ([2, 3, 11, 26]). *Let $\langle U, \pi \rangle$ be an approximation space, where the partition π consists of blocks U_i , $1 \leq i \leq k$, each having cardinality n_i . The co-entropy $G(\pi)$ of partition π is defined by*

$$G(\pi) = \sum_{i=1}^k \frac{n_i}{n} \log n_i, \text{ where } n = \sum_{i=1}^k n_i. \quad (3)$$

It follows immediately from definition that

$$H(\pi) + G(\pi) = \log n.$$

Contrary to the uncertainty measure H , the co-entropy function G achieves the maximum value $\log n$ when $\pi = \tilde{\pi}$ and the minimum value 0 when $\pi = \hat{\pi}$; moreover, it has been known [11] that for any two partitions π and σ of U , if $\sigma < \pi$, then $G(\sigma) < G(\pi)$.

As argued in [2, 3], the entropy $H(\pi)$ can be interpreted as the uncertainty measure of the partition π , while the co-entropy $G(\pi)$ can be regarded as the granularity measure of π . In [21], Sen and Pal introduced two other entropy measures for crisp sets and fuzzy sets with (crisp or fuzzy) equivalence relations or (crisp or fuzzy) tolerance relations, which are based upon the roughness measures of X and of the complement of X in the universe and have been used to analyze the grayness and spatial ambiguities in images. Under the same name, there are some different concepts of entropy in the literature of rough set theory (see, for example, [9, 20]).

3. A novel pair of entropy and co-entropy

In this section, we first introduce a novel entropy and the corresponding co-entropy and then explore their properties.

Let us begin with some notations. Throughout this section, we write $\langle U, \pi \rangle$ for an approximation space and assume that $|U| = n$. Given a $\langle U, \pi \rangle$, we use $\mathcal{A}(U, \pi)$ to denote the set of rough set approximations of all subsets of U . More formally, we set

$$\mathcal{A}(U, \pi) = \left\{ \left(\underline{app}_\pi X, \overline{app}_\pi X \right) \mid X \subseteq U \right\}. \quad (4)$$

It follows from Eq. (4) that $\mathcal{A}(U, \pi)$ has at least two elements: (\emptyset, \emptyset) and (U, U) . If $n = 1$, then $\mathcal{A}(U, \pi)$ exactly consists of the two elements; if $n > 1$ and $\pi = \tilde{\pi}$, then $\mathcal{A}(U, \pi)$ contains one more element (\emptyset, U) ; for any $n \geq 1$, if $\pi = \hat{\pi}$, then we see that $\mathcal{A}(U, \pi) = \{(X, X) \mid X \subseteq U\}$, which consists of 2^n elements. Note that the set $\mathcal{A}(U, \pi)$ is not a multiset, that is, the same element cannot appear more than once in $\mathcal{A}(U, \pi)$. In general, we have that $|\mathcal{A}(U, \pi)| \leq 2^n$ since the subset X of U in Eq. (4) has only 2^n alternatives.

For simplicity, we use m to stand for $|\mathcal{A}(U, \pi)|$. For any $(A_i, A'_i) \in \mathcal{A}(U, \pi)$, $1 \leq i \leq m$, we set

$$\mathcal{A}_i = \left\{ X \subseteq U \mid \left(\underline{app}_\pi X, \overline{app}_\pi X \right) = (A_i, A'_i) \right\} \text{ and } |\mathcal{A}_i| = r_i. \quad (5)$$

In other words, r_i is the number of subsets of U that have the rough set approximation (A_i, A'_i) . It turns out that $\{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_m\}$ gives rise to a partition of $\mathcal{P}(U)$. Therefore, we get by Eq. (4) that

$$\sum_{i=1}^m r_i = 2^n.$$

To illustrate the above concepts, let us see an example.

Example 3.1. Consider $U = \{1, 2, 3, 4\}$ and $\pi = \{\{1, 2\}, \{3, 4\}\}$. In this case, U has 16 subsets. For each subset X of U , we compute the rough set approximation of X ; the results are listed in Table 1.

Hence, we see that

$$\mathcal{A}(U, \pi) = \{(\emptyset, \emptyset), (\emptyset, \{1, 2\}), (\emptyset, \{3, 4\}), (\{1, 2\}, \{1, 2\}), (\emptyset, U), (\{3, 4\}, \{3, 4\}), (\{1, 2\}, U), (\{3, 4\}, U), (U, U)\}.$$

As an example, let us calculate r_2 . By definition,

$$\begin{aligned} r_2 &= \left| \left\{ X \subseteq U \mid \left(\underline{app}_\pi X, \overline{app}_\pi X \right) = (\emptyset, \{1, 2\}) \right\} \right| \\ &= |\{\{1\}, \{2\}\}| \\ &= 2. \end{aligned}$$

This is exactly the number of subsets of U that have the rough set approximation $(\emptyset, \{1, 2\})$, which can be counted from the table. In light of this, we may get Table 2 by rearranging Table 1. It follows immediately from Table 2 that $r_1 = r_4 = r_6 = r_9 = 1$, $r_2 = r_3 = r_7 = r_8 = 2$, and $r_5 = 4$.

Table 1: The subsets and corresponding rough set approximations in Example 3.1.

subset	approximation	subset	approximation	subset	approximation	subset	approximation
\emptyset	(\emptyset, \emptyset)	$\{1\}$	$(\emptyset, \{1, 2\})$	$\{2\}$	$(\emptyset, \{1, 2\})$	$\{3\}$	$(\emptyset, \{3, 4\})$
$\{4\}$	$(\emptyset, \{3, 4\})$	$\{1, 2\}$	$(\{1, 2\}, \{1, 2\})$	$\{1, 3\}$	(\emptyset, U)	$\{1, 4\}$	(\emptyset, U)
$\{2, 3\}$	(\emptyset, U)	$\{2, 4\}$	(\emptyset, U)	$\{3, 4\}$	$(\{3, 4\}, \{3, 4\})$	$\{1, 2, 3\}$	$(\{1, 2\}, U)$
$\{1, 2, 4\}$	$(\{1, 2\}, U)$	$\{1, 3, 4\}$	$(\{3, 4\}, U)$	$\{2, 3, 4\}$	$(\{3, 4\}, U)$	U	(U, U)

Table 2: The rough set approximations and corresponding subsets in Example 3.1.

approximation	subsets	approximation	subsets	approximation	subsets
(\emptyset, \emptyset)	\emptyset	$(\emptyset, \{1, 2\})$	$\{1\}, \{2\}$	$(\emptyset, \{3, 4\})$	$\{3\}, \{4\}$
$(\{1, 2\}, \{1, 2\})$	$\{1, 2\}$	(\emptyset, U)	$\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}$	$(\{3, 4\}, \{3, 4\})$	$\{3, 4\}$
$(\{1, 2\}, U)$	$\{1, 2, 3\}, \{1, 2, 4\}$	$(\{3, 4\}, U)$	$\{1, 3, 4\}, \{2, 3, 4\}$	(U, U)	U

Because we are concerned with the partition granulation of $\langle U, \pi \rangle$ with respect to the approximation operators \underline{app} and \overline{app} , we may assume that every subset of U appears with the same probability $1/2^n$. As a result, the rough set approximation (A_i, A'_i) appears with the accumulative probability $r_i/2^n$ and we thus obtain a probability distribution

$$P(\pi) = \left(\frac{r_1}{2^n}, \frac{r_2}{2^n}, \dots, \frac{r_m}{2^n} \right). \quad (6)$$

According to Shannon's information theory [22], the Shannon entropy function of the probability distribution $P(\pi)$ is defined as follows.

Definition 3.1. *Keep the notations as above. The information entropy $\mathcal{H}(\pi)$ of $\langle U, \pi \rangle$ (with respect to the approximation operators \underline{app} and \overline{app}) is defined by*

$$\mathcal{H}(\pi) = \mathcal{H}(P(\pi)) = - \sum_{i=1}^m \frac{r_i}{2^n} \log \frac{r_i}{2^n}. \quad (7)$$

In the above definition, for simplicity we have used the notation $\mathcal{H}(\pi)$ instead of $\mathcal{H}(U, \pi)$. Following the explanation of Shannon entropy in information theory, the quantity $\mathcal{H}(\pi)$ measures the uncertainty associated to the partition π with respect to the approximation operators \underline{app} and \overline{app} . For instance, the probability distribution corresponding to the partition $\pi = \{\{1, 2\}, \{3, 4\}\}$ in Example 3.1 is

$$P(\pi) = \left(\frac{1}{2^4}, \frac{2}{2^4}, \frac{2}{2^4}, \frac{1}{2^4}, \frac{4}{2^4}, \frac{1}{2^4}, \frac{2}{2^4}, \frac{2}{2^4}, \frac{1}{2^4} \right).$$

It follows from Definition 3.1 that

$$\begin{aligned}
 \mathcal{H}(\pi) &= - \sum_{i=1}^9 \frac{r_i}{2^4} \log \frac{r_i}{2^4} \\
 &= - \left[\frac{1}{2^4} \log \frac{1}{2^4} + \frac{2}{2^4} \log \frac{2}{2^4} + \frac{2}{2^4} \log \frac{2}{2^4} + \frac{1}{2^4} \log \frac{1}{2^4} + \frac{4}{2^4} \log \frac{4}{2^4} \right. \\
 &\quad \left. + \frac{1}{2^4} \log \frac{1}{2^4} + \frac{2}{2^4} \log \frac{2}{2^4} + \frac{2}{2^4} \log \frac{2}{2^4} + \frac{1}{2^4} \log \frac{1}{2^4} \right] \\
 &= 3.
 \end{aligned}$$

Similar to other entropy functions in rough set theory, the information entropy in Definition 3.1 has the following properties.

Theorem 3.1.

- (1) For any $\pi, \sigma \in \Pi(U)$, if $\sigma < \pi$, then $\mathcal{H}(\sigma) > \mathcal{H}(\pi)$.
- (2) The entropy function \mathcal{H} reaches the maximum value n for the finest partition $\hat{\pi}$.
- (3) The entropy function \mathcal{H} reaches the minimum value $n - \frac{2^n-2}{2^n} \log(2^n - 2)$ for the coarsest partition $\check{\pi}$.

Proof. (1) Without loss of generality, we may assume that $\pi = \{U_1, U_2, \dots, U_k\}$ and $\sigma = \{U_a, U_b, U_2, \dots, U_k\}$, where $U_a \cup U_b = U_1$. Suppose that $|\mathcal{A}(U, \pi)| = m$ and for any $(A_i, A'_i) \in \mathcal{A}(U, \pi)$, $1 \leq i \leq m$, we write r_i for

$$\left| \left\{ X \subseteq U \mid (\underline{app}_\pi X, \overline{app}_\pi X) = (A_i, A'_i) \right\} \right|.$$

Based on the partition π , the power set $\mathcal{P}(U)$ is partitioned into m blocks and the i -th block has the cardinality r_i . Similarly, we denote by s_j the cardinality of the j -th block of $\mathcal{P}(U)$ associated to the partition σ . We now consider the elements of $\mathcal{A}(U, \sigma)$. For any $(B_j, B'_j) \in \mathcal{A}(U, \sigma)$, there are two possibilities: One is that $(B_j, B'_j) \in \mathcal{A}(U, \pi)$, say $(B_j, B'_j) = (A_{i_j}, A'_{i_j})$ for some i_j . In this case, it is clear that $s_j = r_{i_j}$. The other case is that $(B_j, B'_j) \in \mathcal{A}(U, \sigma) \setminus \mathcal{A}(U, \pi)$, where the symbol $A \setminus B$ denotes the set of all elements which are members of A but not members of B . It follows that for some i_j ,

$$\left\{ X \subseteq U \mid (\underline{app}_\sigma X, \overline{app}_\sigma X) = (B_j, B'_j) \right\} \subsetneq \left\{ X \subseteq U \mid (\underline{app}_\pi X, \overline{app}_\pi X) = (A_{i_j}, A'_{i_j}) \right\},$$

because the partition σ is strictly finer than π . In this case, we also see that the i_j -th block provided by π is partitioned into smaller blocks and thus $r_{i_j} = \sum_j s_j > s_j$. In summary, we get that either $r_i = s_j$ or $r_i = \sum_j s_{i_j} > s_{i_j}$, and moreover, the latter case must exist as $\sigma < \pi$. We thus assume that $r_i = s_{i_j}$ for $i \in I_1$ and $r_i = \sum_j s_{i_j} > s_{i_j}$ for $i \in I_2$, where $I_2 \neq \emptyset$ and $I_1 \cup I_2 = \{1, 2, \dots, m\}$. Let us compare $\mathcal{H}(\sigma)$ with $\mathcal{H}(\pi)$.

$$\begin{aligned} \mathcal{H}(\pi) &= - \sum_{i=1}^m \frac{r_i}{2^n} \log \frac{r_i}{2^n} \\ &= - \sum_{i \in I_1} \frac{r_i}{2^n} \log \frac{r_i}{2^n} - \sum_{i \in I_2} \frac{r_i}{2^n} \log \frac{r_i}{2^n} \\ &= - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \sum_{i \in I_2} \frac{\sum_j s_{i_j}}{2^n} \log \frac{\sum_j s_{i_j}}{2^n} \\ &= - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \frac{1}{2^n} \sum_{i \in I_2} \left(\sum_j s_{i_j} \right) \left[\log \left(\sum_j s_{i_j} \right) - n \right] \\ &= - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \frac{1}{2^n} \sum_{i \in I_2} \left[\log \left(\sum_j s_{i_j} \right)^{\left(\sum_j s_{i_j} \right)} - n \left(\sum_j s_{i_j} \right) \right] \\ &< - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \frac{1}{2^n} \sum_{i \in I_2} \left[\log \left(\prod_j s_{i_j}^{s_{i_j}} \right) - n \left(\sum_j s_{i_j} \right) \right] \\ &= - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \frac{1}{2^n} \sum_{i \in I_2} \left[\sum_j s_{i_j} \log s_{i_j} - n \left(\sum_j s_{i_j} \right) \right] \\ &= - \sum_{i \in I_1} \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} - \sum_{i \in I_2} \sum_j \frac{s_{i_j}}{2^n} \log \frac{s_{i_j}}{2^n} \\ &= \mathcal{H}(\sigma), \end{aligned}$$

namely, $\mathcal{H}(\sigma) > \mathcal{H}(\pi)$. Therefore, the clause (1) holds.

(2) It follows from (1) that \mathcal{H} reaches the maximum value when $\pi = \hat{\pi}$. In this case, we get by definition that

$$\mathcal{H}(\hat{\pi}) = - \sum_{i=1}^{2^n} \frac{1}{2^n} \log \frac{1}{2^n} = n.$$

This proves (2).

(3) By (1), we see that \mathcal{H} reaches the minimum value when $\pi = \check{\pi}$. In this case, the empty subset \emptyset of U has the rough set approximation (\emptyset, \emptyset) and U itself has the rough set approximation (U, U) . For any proper subset of U , if any, it has the rough set approximation (\emptyset, U) . Hence, $r_1 = r_2 = 1$ and $r_3 = 2^n - 2$. We thus obtain by definition that

$$\begin{aligned} \mathcal{H}(\check{\pi}) &= -\frac{1}{2^n} \log \frac{1}{2^n} - \frac{1}{2^n} \log \frac{1}{2^n} - \frac{2^n - 2}{2^n} \log \frac{2^n - 2}{2^n} \\ &= n - \frac{2^n - 2}{2^n} \log(2^n - 2). \end{aligned}$$

Whence, (3) holds, finishing the proof of the proposition. \square

Note that in the clause (3) of Theorem 3.1, if $n = 1$, the value of the corresponding summand $0 \log 0$ is taken to be 0, which is consistent with the limit:

$$\lim_{x \rightarrow 0^+} x \log x = 0.$$

For later need, let us recall the following definition from [31].

Definition 3.2. Let $\langle U, \pi \rangle$ and $\langle V, \sigma \rangle$ be two approximation spaces, and suppose that $f : U \rightarrow V$ is a mapping.

- (1) The mapping f is called a *homomorphism* from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$ if for any $C \in \pi$, there exists $D \in \sigma$ such that $f(C) \subseteq D$, where $f(C) = \{f(u) \mid u \in C\}$.
- (2) A homomorphism f is called a *monomorphism* if f is an injective mapping.
- (3) A monomorphism f is called *strictly monomorphic* if either there exist $C \in \pi$ and $D \in \sigma$ such that $f(C) \subsetneq D$, namely, $f(C) \subseteq D$ and $f(C) \neq D$, or $|V| > |U|$.
- (4) The mapping f is called an *isomorphism* if the mapping $f : U \rightarrow V$ is bijective, and moreover, both f and its inverse mapping f^{-1} are homomorphisms.

We can now state the following facts.

Proposition 3.1. Let $\langle U, \pi \rangle$ and $\langle V, \sigma \rangle$ be two approximation spaces with $|U| = |V|$, and let $f : U \rightarrow V$ be a mapping.

- (1) If f is a monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, in particular, $\pi \leq \sigma$, then $\mathcal{H}(\pi) \geq \mathcal{H}(\sigma)$.
- (2) If f is a strict monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, in particular, $\pi < \sigma$, then $\mathcal{H}(\pi) > \mathcal{H}(\sigma)$.
- (3) If f is an isomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, then $\mathcal{H}(\pi) = \mathcal{H}(\sigma)$.

Proof. It follows immediately from Definition 3.1 and Theorem 3.1. \square

To measure the granularity with respect to the approximation operators \underline{app} and \overline{app} carried by the partition π , we introduce the concept of co-entropy, which corresponds to the information entropy in Definition 3.1.

Definition 3.3. Keep the notations as in Definition 3.1. The co-entropy $\mathcal{G}(\pi)$ of $\langle U, \pi \rangle$ (with respect to the approximation operators \underline{app} and \overline{app}) is defined by

$$\mathcal{G}(\pi) = \mathcal{G}(P(\pi)) = \sum_{i=1}^m \frac{r_i}{2^n} \log r_i. \quad (8)$$

The quantity $\mathcal{G}(\pi)$ furnishes a measure of the average granularity carried by the partition π as a whole. It follows immediately from definition that

$$\mathcal{H}(\pi) + \mathcal{G}(\pi) = n. \quad (9)$$

It means that the two measures complement each other with respect to the constant quantity $n = |U|$, which is invariant with respect to the choice of the partition π of U .

The co-entropy function \mathcal{G} is of the following properties.

Theorem 3.2.

- (1) For any $\pi, \sigma \in \Pi(U)$, if $\sigma < \pi$, then $\mathcal{G}(\sigma) < \mathcal{G}(\pi)$.
- (2) The co-entropy function \mathcal{G} reaches the minimum value 0 for the finest partition $\hat{\pi}$.
- (3) The co-entropy function \mathcal{G} reaches the maximum value $\frac{2^n-2}{2^n} \log(2^n - 2)$ for the coarsest partition $\check{\pi}$.

Proof. All the clauses follow directly from Theorem 3.1 and Eq. (9). \square

Similar to Proposition 3.1, we have the following observation.

Proposition 3.2. Let $\langle U, \pi \rangle$ and $\langle U, \sigma \rangle$ be two approximation spaces with $|U| = |V|$, and let $f : U \rightarrow V$ be a mapping.

- (1) If f is a monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, in particular, $\pi \leq \sigma$, then $\mathcal{G}(\pi) \leq \mathcal{G}(\sigma)$.
- (2) If f is a strict monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, in particular, $\pi < \sigma$, then $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$.
- (3) If f is an isomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, then $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$.

Proof. It follows immediately from Proposition 3.1 and Eq. (9). \square

As a corollary of Theorem 3.2 and Proposition 3.2, we see that \mathcal{G} is a partition measure on U in the sense of [31, Definition 3.4], that is, \mathcal{G} is nonnegative and satisfies the following two conditions: $\mathcal{G}(\sigma) < \mathcal{G}(\pi)$ if $\sigma < \pi$; $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$ if there exists an isomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$.

Note that our information entropy and co-entropy are not directly based on the blocks of a partition. Therefore, in general they do not satisfy the definition of expected granularity proposed in [28].

4. Relationship of co-entropies between different universes

In the last section, we have seen that if f is a strict monomorphism from $\langle U, \pi \rangle$ to $\langle U, \sigma \rangle$, in particular, $\pi < \sigma$, then $\mathcal{H}(\pi) > \mathcal{H}(\sigma)$ and $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$. In this section, we consider the monotonicities of \mathcal{H} and \mathcal{G} for different universes. In other words, we compare $\mathcal{H}(\pi)$ with $\mathcal{H}(\sigma)$ and $\mathcal{G}(\pi)$ with $\mathcal{G}(\sigma)$ when there exists a strict monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$, where $|V| > |U|$. For convenience, we write $\langle U, \pi \rangle \hookrightarrow \langle V, \sigma \rangle$ if $|V| > |U|$ and there exists a strict monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$.

We start with the following observation on the entropy function H and the co-entropy function G reviewed in Section 2.2. Consider $\langle U_1, \pi_1 \rangle = \langle \{1\}, \{\{1\}\} \rangle$, $\langle U_2, \pi_2 \rangle = \langle \{1, 2\}, \{\{1\}, \{2\}\} \rangle$, and $\langle U_3, \pi_3 \rangle = \langle \{1, 2, 3\}, \{\{1, 3\}, \{2\}\} \rangle$. Clearly,

$$\langle U_1, \pi_1 \rangle \hookrightarrow \langle U_2, \pi_2 \rangle \hookrightarrow \langle U_3, \pi_3 \rangle.$$

It is easy to check by Definition 2.1 that $H(\pi_1) = 0$, $H(\pi_2) = 1$, and $H(\pi_3) = \log 3 - \frac{2}{3} < 1$. This means that the entropy function H is not monotonic. By the way, we can get by a direct computation that $\mathcal{G}(\pi_1) = 0$, $\mathcal{G}(\pi_2) = 0$, and $\mathcal{G}(\pi_3) = \frac{1}{2}$.

Let us continue to discuss the monotonicity of co-entropy function G . Consider $\langle U_1, \pi_1 \rangle = \langle \{1\}, \{\{1\}\} \rangle$, $\langle U_2, \pi_2 \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$, and $\langle U_3, \pi_3 \rangle = \langle \{1, 2, 3\}, \{\{1, 2\}, \{3\}\} \rangle$. Again, we see that

$$\langle U_1, \pi_1 \rangle \hookrightarrow \langle U_2, \pi_2 \rangle \hookrightarrow \langle U_3, \pi_3 \rangle.$$

It is easy to check by Definition 2.2 that $G(\pi_1) = 0$, $G(\pi_2) = 1$, and $G(\pi_3) = \frac{2}{3}$. This shows that the co-entropy function G is not monotonic either. In this case, we can obtain by a direct computation that $\mathcal{G}(\pi_1) = 0$, $\mathcal{G}(\pi_2) = \frac{1}{2}$, and $\mathcal{G}(\pi_3) = \frac{1}{2}$.

Finally, we address the monotonicity of entropy function \mathcal{H} . Consider $\langle U_1, \pi_1 \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$, $\langle U_2, \pi_2 \rangle = \langle \{1, 2, 3\}, \{\{1, 2\}, \{3\}\} \rangle$, and $\langle U_3, \pi_3 \rangle = \langle \{1, 2, 3, 4\}, \{\{1, 2, 4\}, \{3\}\} \rangle$. Obviously, we have that

$$\langle U_1, \pi_1 \rangle \hookrightarrow \langle U_2, \pi_2 \rangle \hookrightarrow \langle U_3, \pi_3 \rangle.$$

By a routine computation we can get that $\mathcal{H}(\pi_1) = \frac{3}{2}$, $\mathcal{H}(\pi_2) = \frac{5}{2}$, and $\mathcal{H}(\pi_3) = \frac{13}{4} - \frac{3}{4} \log 3 < \frac{5}{2}$. Consequently, the entropy function \mathcal{H} is not monotonic either. On the other hand, it follows from Eq. (9) that $\mathcal{G}(\pi_1) = \frac{1}{2}$, $\mathcal{G}(\pi_2) = \frac{1}{2}$, and $\mathcal{G}(\pi_3) = \frac{3}{4} + \frac{3}{4} \log 3$.

As a result, in all the above three cases we always have that

$$\mathcal{G}(\pi_1) \leq \mathcal{G}(\pi_2) \leq \mathcal{G}(\pi_3).$$

We thus conjecture that $\mathcal{G}(\pi) \leq \mathcal{G}(\sigma)$ whenever $\langle U, \pi \rangle \hookrightarrow \langle V, \sigma \rangle$. Indeed, it holds true, as we will see later.

To prove the conjecture, it is convenient to introduce the following notion and a key lemma.

Definition 4.1. Let $\langle U, \pi \rangle$ be an approximation space and $a \notin U$. The approximation space $\langle U \cup \{a\}, \pi \cup \{\{a\}\} \rangle$ is called the one-point extension of $\langle U, \pi \rangle$ by a . We say that $\langle V, \sigma \rangle$ is a one-point extension of $\langle U, \pi \rangle$ if $\langle V, \sigma \rangle = \langle U \cup \{a\}, \pi \cup \{\{a\}\} \rangle$ for some a .

For example, $\langle U_2, \pi_2 \rangle = \langle \{1, 2, 3\}, \{\{1, 2\}, \{3\}\} \rangle$ is the one-point extension of $\langle U_1, \pi_1 \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$ by 3.

The following lemma shows that one-point extension does not change co-entropy.

Lemma 4.1. Let $\langle V, \sigma \rangle$ be a one-point extension of $\langle U, \pi \rangle$. Then $\mathcal{G}(\sigma) = \mathcal{G}(\pi)$.

Proof. Suppose that $\pi = \{U_1, U_2, \dots, U_k\}$ and $\sigma = \{U_1, U_2, \dots, U_k, \{a\}\}$, where $a \notin U$; assume that $\mathcal{A}(U, \pi) = \{(A_i, A'_i) \mid 1 \leq i \leq m\}$ and $\mathcal{A}_i = \left\{ X \subseteq U \mid (\underline{app}_\pi X, \overline{app}_\pi X) = (A_i, A'_i) \right\}$ with $|\mathcal{A}_i| = r_i$. It thus follows that

$$\mathcal{A}(V, \sigma) = \mathcal{A}(U, \pi) \cup \{(A_i \cup \{a\}, A'_i \cup \{a\}) \mid 1 \leq i \leq m\}.$$

For any $(B_i, B'_i) \in \mathcal{A}(V, \sigma)$, we write \mathcal{B}_i for $\left\{ X \subseteq V \mid (\underline{app}_\sigma X, \overline{app}_\sigma X) = (B_i, B'_i) \right\}$ and s_i for $|\mathcal{B}_i|$. If $(B_i, B'_i) = (A_i, A'_i) \in \mathcal{A}(U, \pi)$, then we see that $\mathcal{B}_i = \mathcal{A}_i$ and thus $s_i = r_i$ in this case. If $(B_i, B'_i) = (A_i \cup \{a\}, A'_i \cup \{a\}) \in \{(A_i \cup \{a\}, A'_i \cup \{a\}) \mid 1 \leq i \leq m\}$, then we have that $\mathcal{B}_i = \{X \cup \{a\} \mid X \in \mathcal{A}_i\}$ and $s_i = r_i$ still holds in this case. Therefore, we get by Definition 3.3 that

$$\begin{aligned} \mathcal{G}(\sigma) &= \sum_{i=1}^m \frac{r_i}{2^{n+1}} \log r_i + \sum_{i=1}^m \frac{r_i}{2^{n+1}} \log r_i \\ &= \sum_{i=1}^m \frac{r_i}{2^n} \log r_i \\ &= \mathcal{G}(\pi), \end{aligned}$$

finishing the proof of the lemma. □

For subsequent need, we would like to generalize Definition 4.1 as follows.

Definition 4.2. Let $\langle U, \pi \rangle$ and $\langle V, \sigma \rangle$ be two approximation spaces. We say that $\langle V, \sigma \rangle$ is a multi-one-point extension of $\langle U, \pi \rangle$ if there are approximation spaces $\langle U_i, \pi_i \rangle$, $0 \leq i \leq l$, with $\langle U_0, \pi_0 \rangle = \langle U, \pi \rangle$ and $\langle U_l, \pi_l \rangle = \langle V, \sigma \rangle$ such that each $\langle U_i, \pi_i \rangle$, $1 \leq i \leq l$, is a one-point extension of $\langle U_{i-1}, \pi_{i-1} \rangle$.

For example, $\langle V, \sigma \rangle = \langle \{1, 2, 3, 4\}, \{\{1, 2\}, \{3\}, \{4\}\} \rangle$ is a multi-one-point extension of $\langle U, \pi \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$. In fact, we may take $\langle U_0, \pi_0 \rangle = \langle U, \pi \rangle$, $\langle U_1, \pi_1 \rangle = \langle \{1, 2, 3\}, \{\{1, 2\}, \{3\}\} \rangle$, and $\langle U_2, \pi_2 \rangle = \langle V, \sigma \rangle$.

The following fact follows immediately from Lemma 4.1.

Corollary 4.1. If $\langle V, \sigma \rangle$ is a multi-one-point extension of $\langle U, \pi \rangle$, then $\mathcal{G}(\sigma) = \mathcal{G}(\pi)$.

In light of the above corollary, let us refer to multi-one-point extensions as one-point extensions for simplicity. Further, we have the following observation.

Theorem 4.1. Suppose that there is a monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$. If there exists $\langle U', \pi' \rangle$ that satisfies the following two conditions:

- (a) either $\langle U', \pi' \rangle = \langle U, \pi \rangle$ or $\langle U', \pi' \rangle$ is a one-point extension of $\langle U, \pi \rangle$,
- (b) $\langle U', \pi' \rangle$ is isomorphic to $\langle V, \sigma \rangle$,

then $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$; otherwise, $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$.

Proof. We first consider the case that there exists $\langle U', \pi' \rangle$ that satisfies the conditions (a) and (b). In this case, if $\langle U', \pi' \rangle = \langle U, \pi \rangle$ and $\langle U', \pi' \rangle$ is isomorphic to $\langle V, \sigma \rangle$, then $|V| = |U|$ and we see by Proposition 3.2 that $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$. If $\langle U', \pi' \rangle$ is a one-point extension of $\langle U, \pi \rangle$ and $\langle U', \pi' \rangle$ is isomorphic to $\langle V, \sigma \rangle$, then we get that $\mathcal{G}(\pi) = \mathcal{G}(\pi')$ by Corollary 4.1 and $\mathcal{G}(\pi') = \mathcal{G}(\sigma)$ by Proposition 3.2. Consequently, $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$.

We now consider the case that there does not exist $\langle U', \pi' \rangle$ such that the conditions are satisfied. It forces that the monomorphism, say f , from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$ is strict. Two cases need to consider. One is that $|V| = |U|$. In this case, it follows from Proposition 3.2 that $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$. The other case is that $|V| > |U|$. In this case, let us set $\langle V_1, \sigma_1 \rangle = \langle f(U), f(\pi) \rangle$, where $f(U)$ is the image of U under f and $f(\pi) = \{f(U') \mid U' \in \pi\}$. In fact, f gives rise to an isomorphism between $\langle U, \pi \rangle$ and $\langle V_1, \sigma_1 \rangle$. Therefore, $\mathcal{G}(\pi) = \mathcal{G}(\sigma_1)$. Note that $V_1 = f(U) \subseteq V$. We now take $\langle V_2, \sigma_2 \rangle$ as follows:

$$V_2 = V, \sigma_2 = \sigma_1 \cup \{\{a\} \mid a \in V \setminus V_1\}.$$

It follows that $\langle V_2, \sigma_2 \rangle$ is a one-point extension of $\langle V_1, \sigma_1 \rangle$. Hence, $\mathcal{G}(\sigma_1) = \mathcal{G}(\sigma_2)$. Because f is a strict monomorphism, we see that $\langle U, \pi \rangle \hookrightarrow \langle V_2, \sigma_2 \rangle$ and $\sigma_2 < \sigma$. Whence, we get $\mathcal{G}(\sigma_2) < \mathcal{G}(\sigma)$ by Theorem 3.2. As a result, $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$. This completes the proof of the theorem. \square

Let us provide an informal explanation of Theorem 4.1. The hypothesis that there is a monomorphism from $\langle U, \pi \rangle$ to $\langle V, \sigma \rangle$ means that $\langle U, \pi \rangle$ is finer than $\langle V, \sigma \rangle$. In the special case that the monomorphism is not strict, we have that $\langle U, \pi \rangle$ and $\langle V, \sigma \rangle$ are isomorphic, and thus, they have the same co-entropy. If the monomorphism is strict, then after renaming the elements of U , we can get a finer partition than $\langle V, \sigma \rangle$ by using one-point extensions. Theorem 4.1 says that $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$ if π is finer than σ .

We end this section with several examples.

Example 4.1. A trivial example is that $\langle U, \pi \rangle = \langle \{1, 2, 3\}, \{\{1, 2\}, \{3\}\} \rangle$ and $\langle V, \sigma \rangle = \langle \{a, b, c\}, \{\{a, b\}, \{c\}\} \rangle$. The mapping f that maps 1, 2, and 3 to a , b , and c respectively is a monomorphism. In fact, f is an isomorphism. Hence, $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$. A direct computation shows that $\mathcal{G}(\pi) = 0.5 = \mathcal{G}(\sigma)$.

Consider $\langle U, \pi \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$ and $\langle V, \sigma \rangle = \langle \{a, b, c, d\}, \{\{a, b\}, \{c\}, \{d\}\} \rangle$. The mapping f that maps 1 and 2 to a and b respectively is a monomorphism, which yields that $\langle U, \pi \rangle$ is isomorphic to $\langle V_1, \sigma_1 \rangle = \langle \{a, b\}, \{\{a, b\}\} \rangle$. Clearly, we can get $\langle V, \sigma \rangle$ by one-point extensions of $\langle V_1, \sigma_1 \rangle$. Therefore, $\mathcal{G}(\pi) = \mathcal{G}(\sigma)$. On the other hand, we can get by a computation that $\mathcal{G}(\pi) = 0.5 = \mathcal{G}(\sigma)$.

Finally, consider $\langle U, \pi \rangle = \langle \{1, 2\}, \{\{1, 2\}\} \rangle$ and $\langle V, \sigma \rangle = \langle \{a, b, c, d\}, \{\{a, b\}, \{c, d\}\} \rangle$. As mentioned earlier, the mapping f that maps 1 and 2 to a and b respectively is a monomorphism, which gives an isomorphism between $\langle U, \pi \rangle$ and $\langle V_1, \sigma_1 \rangle = \langle \{a, b\}, \{\{a, b\}\} \rangle$. We can get $\langle V_2, \sigma_2 \rangle = \langle \{a, b, c, d\}, \{\{a, b\}, \{c\}, \{d\}\} \rangle$ by one-point extensions of $\langle V_1, \sigma_1 \rangle$. Clearly, $\sigma_2 < \sigma$. As a result, $\mathcal{G}(\pi) < \mathcal{G}(\sigma)$. On the other hand, we can obtain by a direct computation that $\mathcal{G}(\pi) = 0.5$ and $\mathcal{G}(\sigma) = 0.75$.

5. Conclusion

In this paper, we have proposed the novel notions of entropy and co-entropy by taking both partitions and the lower and upper approximations into account. Some desirable properties of the entropy and co-entropy have been presented. Furthermore, we have investigated the relationship of co-entropies between different universes.

There are several problems which are worth further studying. Firstly, the present work focuses on the classical rough sets based on partitions. It would be interesting to generalize the notions of entropy and co-entropy here into the framework of covering rough sets [4, 19, 29] or fuzzy rough sets [5]. It is also interesting to compare the entropies (co-entropies) under some special homomorphisms such as neighborhood-consistent functions introduced in [33]. Secondly, it remains to develop the corresponding roughness measure based on the entropy or co-entropy for measuring numerically the roughness of an approximation. Finally, the conditioned entropy and conditioned co-entropy [2] are yet to be addressed in our framework.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 60821001, 60873191, 60903152, and 61070251.

References

- [1] Beaubouef, T., Petry, F. E., Arora, G., 1998. Information-theoretic measures of uncertainty for rough sets and rough relational databases. *Inform. Sci.* 109, 185–195.
- [2] Bianucci, D., Cattaneo, G., 2009. Information entropy and granulation co-entropy of partitions and coverings: A summary. In: Peters, J., Skowron, A., Wolski, M., Chakraborty, M., Wu, W.-Z. (Eds.), *Transactions on Rough Sets X*. Vol. 5656 of *Lect. Notes Comput. Sci.* Springer Berlin/Heidelberg, pp. 15–66.
- [3] Bianucci, D., Cattaneo, G., Ciucci, D., 2007. Entropies and co-entropies of coverings with application to incomplete information systems. *Fund. Inform.* 75, 77–105.
- [4] Bryniarski, E., 1989. A calculus of rough sets of the first order. *Bull. Pol. Acad. Sci.* 36 (16), 71–77.
- [5] Dubois, D., Prade, H., 1990. Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* 17, 191–208.
- [6] Düntsch, I., Gediga, G., 1998. Uncertainty measures of rough set prediction. *Artif. Intell.* 106 (1), 109–137.
- [7] Hartley, R. V. L., 1928. Transmission of information. *Bell Syst. Tech. J.* 7, 535–564.
- [8] Liang, J., Shi, Z., Li, D., Wierman, M. J., 2006. Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen. Syst.* 35 (6), 641–654.
- [9] Liang, J. Y., Chin, K. S., Dang, C. Y., Yam, R. C. M., 2002. A new method for measuring uncertainty and fuzziness in rough set theory. *Int. J. Gen. Syst.* 31, 331–342.
- [10] Liang, J. Y., Qian, Y. H., 2008. Information granules and entropy theory in information systems. *Sci. China Ser. F: Inform. Sci.* 51 (10), 1427–1444.
- [11] Liang, J. Y., Shi, Z. Z., 2004. The information entropy, rough entropy and knowledge granulation in rough set theory. *Int. J. Uncert. Fuzz. Knowl. Syst.* 12 (1), 37–46.
- [12] Lin, T. Y., Cercone, N. (Eds.), 1997. *Rough Sets and Data Mining*. Kluwer Academic Publishers, Boston.
- [13] Miao, D. Q., Wang, J., 1998. On the relationships between information entropy and roughness of knowledge in rough set theory (in Chinese). *Pattern Recognit. Artif. Intell.* 11 (1), 34–40.
- [14] Pawlak, Z., 1982. Rough sets. *Int. J. Comput. Inform. Sci.* 11 (5), 341–356.
- [15] Pawlak, Z., 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Boston.
- [16] Pawlak, Z., 1992. Rough sets: a new approach to vagueness. In: Zadeh, L. A., Kacprzyk, J. (Eds.), *Fuzzy Logic for the Management of Uncertainty*. John Wiley & Sons, Inc, New York, pp. 105–118.
- [17] Polkowski, L., Skowron, A. (Eds.), 1998. *Rough Sets and Current Trends in Computing*. Vol. 1424. Springer, Berlin.
- [18] Polkowski, L., Skowron, A. (Eds.), 1998. *Rough Sets in Knowledge Discovery*. Vol. 1 and 2. Physica-Verlag, Heidelberg.
- [19] Pomykala, J. A., 1987. Approximation operations in approximation space. *Bull. Pol. Acad. Sci.* 35 (9-10), 653–662.
- [20] Qian, Y. H., Liang, J. Y., 2008. Combination entropy and combination granulation in rough set theory. *Int. J. Uncert. Fuzz. Knowl. Syst.* 16 (2), 179–193.
- [21] Sen, D., Pal, S. K., 2009. Generalized rough sets, entropy, and image ambiguity measures. *IEEE Trans. Syst., Man, Cybern. B, Cybern.* 39 (1), 117–128.
- [22] Shannon, C. E., 1948. A mathematical theory of communication, I, II. *Bell Syst. Techn. J.* 27, 379–423, 623–656.
- [23] Wierman, M., 1999. Measuring uncertainty in rough set theory. *Int. J. Gen. Syst.* 28, 283–297.
- [24] Xu, W. H., Zhang, X. Y., Zhang, W. X., 2009. Knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems. *Appl. Soft Comput.* 9 (4), 1244–1251.
- [25] Yao, Y. Y., 2003. Information-theoretic measures for knowledge discovery and data mining. In: Karmeshu (Ed.), *Entropy Measures, Maximum Entropy and Emerging Applications*. Springer, Berlin, pp. 115–136.
- [26] Yao, Y. Y., 2003. Probabilistic approaches to rough sets. *Expert Syst.* 20 (5), 287–297.
- [27] Yao, Y. Y., 2010. Notes on rough set approximations and associated measures. *J. Zhejiang Ocean Univ. (Natur. Sci.)* 29, 399–410.
- [28] Yao, Y. Y., Zhao, L. Q., 2009. Granularity of partitions, (private communication).
- [29] Zakowski, W., 1983. Approximations in the space (u, π) . *Demonstr. Math.* 16, 761–769.
- [30] Zhong, N., Yao, Y., Ohshima, M., 2003. Peculiarity oriented multidatabase mining. *IEEE Trans. Knowl. Data Eng.* 15 (4), 952–960.
- [31] Zhu, P., 2009. An axiomatic approach to the roughness measure of rough sets. *Fund. Inform.* To appear, available at: Arxiv preprint arXiv:0911.5395.
- [32] Zhu, P., 2009. An improved axiomatic definition of information granulation. Arxiv preprint arXiv:0908.3999.
- [33] Zhu, P., Wen, Q., 2010. Some improved results on communication between information systems. *Inform. Sci.* 180 (18), 3521–3531.
- [34] Zhu, W., Wang, F. Y., 2006. Covering based granular computing for conflict analysis. Vol. 3975 of *Lect. Notes Comput. Sci.* pp. 566–571, *IEEE Int. Conf. Intell. Secur. Inform. (ISI 2006)*, San Diego, CA, May 23–24, 2006.
- [35] Ziarko, W. (Ed.), 1994. *Rough Sets, Fuzzy Sets, and Knowledge Discovery*. Springer-Verlag, Berlin.